New Polynomial-Based Molecular Descriptors with Low Degeneracy

Matthias Dehmer*, Laurin A. J. Mueller, Armin Graber

Institute for Bioinformatics and Translational Research, Private Universität für Medizinische Informatik und Technik (UMIT), Hall in Tirol, Austria

Abstract

In this paper, we introduce a novel graph polynomial called the 'information polynomial' of a graph. This graph polynomial can be derived by using a probability distribution of the vertex set. By using the zeros of the obtained polynomial, we additionally define some novel spectral descriptors. Compared with those based on computing the ordinary characteristic polynomial of a graph, we perform a numerical study using real chemical databases. We obtain that the novel descriptors do have a high discrimination power.

Introduction

The study of specific structural properties of graphs by using algebraic polynomial representations has been a well-known and fruitful concept for several decades [1–6]. In particular, graph polynomials have been either used for describing combinatorial graph invariants or to characterize chemical structures by using the coefficients or the zeros of a graph polynomial [4,7,8]. As outlined by Gutman [4] and Ivanciuc et al. [9], various topics in mathematical chemistry like Hückel-molecular orbital theory, the theory of aromaticity, and the development of topological indices (e.g., Hosoya index etc.) rely on graph polynomials. Indeed, a very important graph polynomial is the characteristic polynomial of a graph which has been intensely studied by Cvetkovic [10] when exploring structural properties of a graph related to its eigenvalues. Several methods to compute the characteristic polynomial explicitly were also developed [9]. Afterwards, various other graph polynomials [8,9,11] such as the Laplacian polynomial, Matching polynomial, Mühlheim polynomial, Distance polynomial and the Wiener Polynomial etc. have been developed for investigating multifaceted aspects of chemical structures.

In this paper, we use the zeros of a novel graph polynomial to derive molecular descriptors. Before outlining the contribution of our paper, we briefly sketch some approaches surveyed by Randić et al. [12] who gave a thorough overview on existing eigenvalue-based descriptors and various applications thereof. An early starting point in this area was initiated by Lovász and Pelikán [13]: They employed the leading positive eigenvalue of the characteristic polynomial as a measure for detecting branching of trees. More advanced concepts to quantify branching of chemical graphs and DNA structures relying on graph-eigenvalues can be also found in Randić's survey [12]. Particularly, the largest and second largest eigenvalues of other graph-theoretical matrices [12], the sum of their positive eigenvalues, the multiplicity of the zero eigenvalue and other spectral indices were also studied [12,14] to check their ability for serving as molecular descriptors. Importantly, numerical results were reported [12] when applying such descriptors to explore correlations of several physico-chemical properties of chemicals. As a final remark, other eigenvalue-based measures and graph polynomials have been used in various scientific disciplines as well. For instance, graph spectra were employed to derive indices for measuring the structural similarity of networks [15]. In social network analysis and biology, such indices turned out to be useful for defining centrality measures [16,17]. Estrada [18] derived an eigenvalue-based measure (called the Estrada index) for examining the degree of folding of proteins. More generally, polynomial-based approaches have also been employed to study structural aspects of networks, e.g., see [19,20].

The main contribution of our paper is as follows: Firstly, we define a novel graph polynomial which we call the *information polynomial* of a graph $G = (V,E), |V| = n$. In our sense, we infer this polynomial by using a probability distribution of length n for deriving a graph-theoretical matrix $M(G)$ (see Equation (8) in the section 'Methods'). Secondly, we use the zeros of this polynomial to derive some novel molecular descriptors. Then, we investigate their uniqueness and compare our results with other descriptors. An illustration of our workflow is shown in Figure 1.

Methods

The Information Polynomial of a Graph

In this section, we define a novel graph polynomial as well as some spectral molecular descriptors by combining information-theoretic and algebraic methods [21,22]. To provide measures (descriptors) for quantifying structural information of a graph
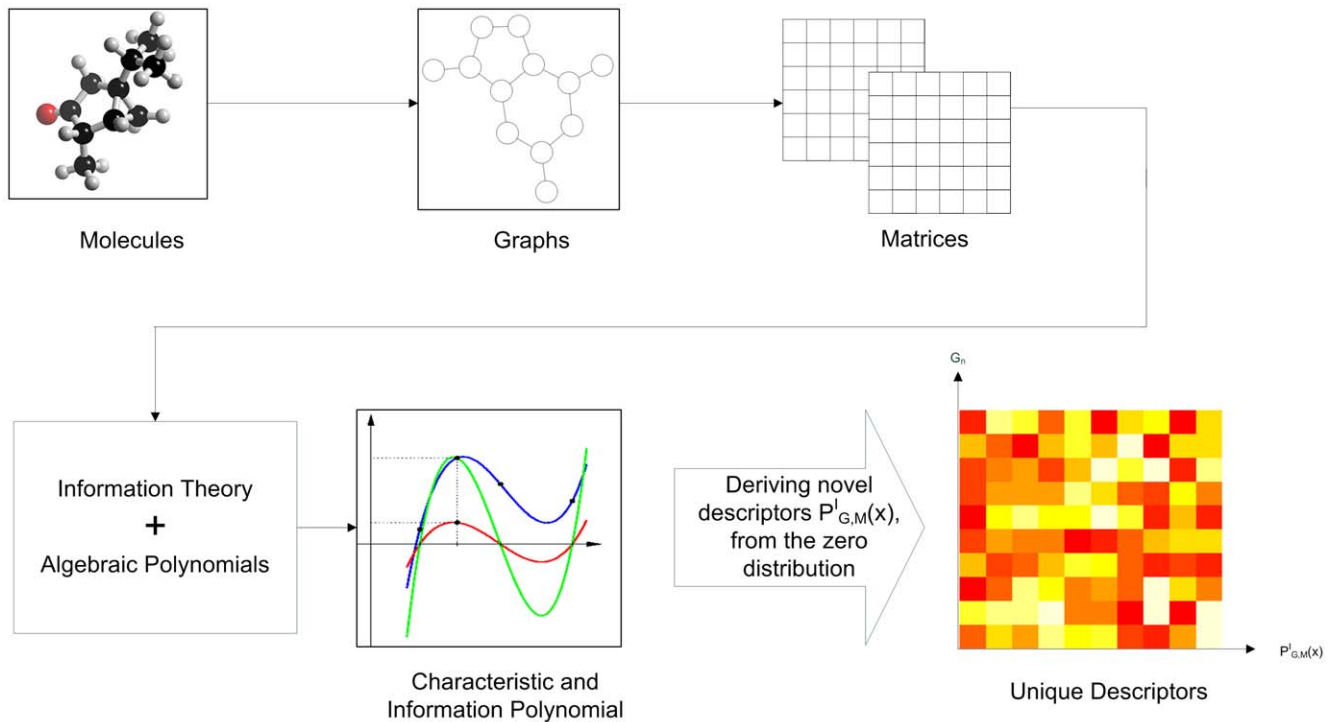
Figure 1. An illustration of the main contribution of this study.
doi:10.1371/journal.pone.0011393.g001

meaningfully, we develop the notation of the 'information polynomial' of a graph. Then, the main idea is to derive a polynomial representation whose coefficients or other characteristics, e.g., its zeros capture structural information of the underlying graph. We want to remark that the expression 'information polynomial' has already been used [23,24] in another context by determining the relative information $I(G)$ defined as the combinatorial entropy of an oriented graph G.

In particular, it was shown [24] that $I(G)$ can be expressed by

$$I(G) = \frac{1}{2} \log(P_G(\lambda)), \tag{1}$$

and

$$I_\sigma(G) = \frac{1}{2} \log\left(\frac{\lambda P_G(\lambda)}{P_G^\sigma(\lambda)}\right), \tag{2}$$

where

$$P_G(\lambda) = \det(a_{\alpha\beta})_{\alpha, \beta \in G}, \tag{3}$$

and

$$P_G^\sigma(\lambda) = \det(a_{\alpha\beta})_{\alpha, \beta \in G \cup \{\sigma\}}. \tag{4}$$

The kind of information polynomial we want to introduce here is notably different from the just mentioned one. The reason why we keep this notation is that the underlying probability distribution as well as the resulting polynomial captures structural information of a graph. In our case, we use an information functional [25–27] leading to vertex probability values that can be used to define a graph-theoretical matrix.

We now start with a probability distribution

$$P_G(V) := (p^f(v_1), p^f(v_2), \ldots, p^f(v_n)), \tag{5}$$

associated to a graph that has already been used to obtain information measures for determining the structural information content of a graph [25–28]. The general procedure is as follows: One assigns a probability value to each vertex of a given graph by using a certain information functional f [25–27] where f represents a function that maps vertices, or more generally graph elements (when using other invariants), to the non-negative reals. Then, by using Shannon's entropy [29], the structural information content will be defined and interpreted as the entropy of the underlying graph topology [26,27].

Let's introduce this framework formally. The quantities serving as vertex probabilities are defined by [26]

$$p^f(v_i) := \frac{f(v_i)}{\sum_{j=1}^n f(v_j)}, \quad \forall v_i \in V. \tag{6}$$

Further, a family of graph entropy measures could be obtained by [26]

$$I_f(G) := -\sum_{i=1}^n \frac{f(v_i)}{\sum_{j=1}^n f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^n f(v_j)}\right). \tag{7}$$

Starting from a graph G and the probability distribution (see Equation (5)), we now define a matrix for computing an algebraic polynomial which we also call the information polynomial of G. In contrast to the briefly sketched polynomial $P_G(\lambda)$ [23,24], our main goal is to investigate the information polynomial (see Definition (2)) for deriving molecular descriptors from the

underlying zero distribution. Further, we remark that our polynomial (see Equation (2)) does not serve as an input for determining a kind of information value. Instead, the structural information of a graph will be captured by the polynomial.

Definition 1. *Let* $G = (V, E)$ *be a graph and let* $P_G(V)$ *be the probability distribution assigned to vertex set of* G. *We define the matrix*

$$M(G) := \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & 1 \end{pmatrix}. \tag{8}$$

We set $a_{ii} = 1, 1 \leq i \leq n$ and

$$a_{ij} = 1 - |p^f(v_i) - p^f(v_j)| \beta(d(v_i, v_j)). \tag{9}$$

Definition 2. *We define* $P_{G,M}^I(x) := \det(M(G) - xE)$. $P_{G,M}^I(x)$ *is called the information polynomial of* G.

We remark that starting from Equation (8), (9), we see immediately that $M(G)$ is symmetric, i.e., $M(G) = (M(G))^T$. Then, it is clear that all eigenvalues of $P_{G,M}^I(x)$ are real [30]. Also, we point out that to comprise the connection between v_i and v_j, we use a function $\beta = \beta(d(v_i, v_j))$ (see Equation (9)) depending on the shortest distance between the corresponding vertices. To calculate $P_{G,M}^I(x)$ for some simple graphs exemplarily, we consider Figure 2 and set

$$f(v_i) = c_1 |S_1(v_i, G)| + c_2 |S_2(v_i, G)| + \cdots + c_{\rho(G)} |S_{\rho(G)}(v_i, G)|, \\ c_k > 0, 1 \leq k \leq \rho(G), \tag{10}$$

and $c_1 = 4, c_2 = 3, c_3 = 2, c_4 = 1$ and $\beta(d(v_i, v_j)) = \frac{1}{2^{d(v_i, v_j)}}$ for determining the probability distribution $P_{G_i}(V)$. $\rho(G)$ is the diameter of G. We yield

$$P_{G_1,M}^I(x) = -x^5 + 5x^4 - 0.2230x^3 + 9.548 \cdot 10^{-5} x^2 \\ + 7.851 \cdot 10^{-5} x - 6.515 \cdot 10^{-7}, \tag{11}$$

$$P_{G_2,M}^I(x) = -x^5 + 5x^4 - 0.4170x^3, \tag{12}$$

$$P_{G_3,M}^I(x) = -x^5 + 5x^4. \tag{13}$$

and the corresponding sets of zeros are

$$Z_{M(G_1)}(x) = \{4.9550, 0.0187, 0.0339, 0.0110, -0.0187\}, \tag{14}$$

Figure 2. Undirected example graphs G_i.
doi:10.1371/journal.pone.0011393.g002

$$Z_{M(G_2)}(x) = \{4.9151, 0.0848, 0, 0, 0\}, \tag{15}$$

$$Z_{M(G_3)}(x) = \{5, 0, 0, 0, 0\}. \tag{16}$$

In particular, we always get
Proposition 1.

$$P_{K_n,M}^I(x) = (-1)^n x^n + (-1)^{n-1} n x^{n-1}. \tag{17}$$

The positive eigenvalue is $x = n$.

Novel Descriptors

In the following, we define some novel descriptors derived from the just introduced polynomial. The idea is to use the underlying zero distribution of $P_{G,M}^I(x)$.

Definition 3. *Let* $P_{G,M}^I(x)$ *be the information polynomial of* G *and let* $\sigma_1, \sigma_2, \ldots, \sigma_n$ *be its real zeros.*

$$I_{M,G}^1 := \sqrt{|\sigma_1|} + \sqrt{|\sigma_2|} + \cdots + \sqrt{|\sigma_n|}, \tag{18}$$

$$I_{M,G}^2 := \sigma = \max(\sigma_1, \sigma_2, \ldots, \sigma_n), \tag{19}$$

$$I_{M,G}^3 := \frac{1}{n} \sum_{i=1}^{n} (\sigma_i - \bar{\sigma})^2, \tag{20}$$

$$I_{M,G}^4 := \sqrt{I_{M,G}^3}. \tag{21}$$

Results and Discussion

The aim of this section is to evaluate the just defined descriptors (see previous section) in terms of their uniqueness (degeneracy) [31,32]. This property of a molecular descriptor relates to the ability to distinguish graphs as uniquely as possible by calculating the underlying graph measure. In general, a descriptor is called degenerated if there are at least two non-isomorphic graphs possessing the same value. For instance, concrete studies [32,33] to explore this problem by using isomeric and lattice structures were performed. Particularly, Bonchev et al. [7] investigated the degeneracy of information-theoretic indices relying on Shannon's entropy. Further, Todeschini et al. [34] evaluated several known non-information-theoretic and information-theoretic topological indices based on a large set of real chemical compounds. Another important problem is to quantify the degree of degeneracy of a given index. For this, a sensitivity measure as well as an information-theoretic measure have been developed by Konstantinova [32,33] and Todeschini et al. [34], respectively.

Datasets

- AG 3982: To create this database, we used the benchmark database called Ames mutagenicity [35,36]. This database has originally been used for predicting the mutagenicity of chemical compounds [36]. Note that the Ames database has been created from six different public sources [35,36]. By

starting from the original database Ames mutagenicity [35,36] containing 6512 chemical compounds, we created AG 3982 by filtering out isomorphic graphs based on the software SubMat [37]. Finally, this procedure resulted in 3982 structurally different skeletons, that is, all atoms and all bonds are considered as equal. It holds $2 \leq n \leq 109 \, \forall \, G \in$ AG 3982.

- MS 2265: To create this dataset, the mass spectral database NIST [38] was used [27]. The entire database contains approximately 100000 chemical structures from organic compounds [27]. A set of 4000 structures with 4 to 19 non-hydrogen atoms (vertices of the graph) have been randomly selected from the database. Finally, 2265 of them have different skeletons (all atoms and all bonds considered to be equal, all hydrogen atoms removed). It holds $4 \leq n \leq 19 \, \forall \, G \in$ MS 2265.

Software and Technical Processing of Graphs

We implemented the novel descriptors and performed all involved calculations by using the programming language R [39] and the freely available library 'graph'. In order to generate the underlying graph structures representing chemical compounds, we used Molfile format [40]. We remark that the graphs originating from AG 3982 were originally available in Smiles format. Thus, we converted them to Molfile format (SDF) using a Python procedure. The graphs from MS 2265 were originally available as Molfiles. In both cases, a R procedure was developed to create the adjacency matrices as well as the matrices $M(G)$ for calculating the novel descriptors from the available Molfiles. To determine their degree of degeneracy (sensitivity measure) in the next section, we use Equation (10) and exponentially decreasing parameters chosen as

$$c_1 := \rho(G), c_2 := \rho(G)e^{-1}, \ldots, c_{\rho(G)} := \rho(G)e^{-\rho(G)+1}. \quad (22)$$

Uniqueness of the Descriptors

We now interpret the results we have obtained from calculating the sensitivity index [32]

$$S(I) = \frac{|\mathcal{G}| - |\mathcal{G}_i|}{|\mathcal{G}|}, \quad (23)$$

of a descriptor I as follows. Here, $|\mathcal{G}|$ is the cardinality of a given set of graphs $\mathcal{G}$ and $|\mathcal{G}_i|$ denotes the number of graphs $\mathcal{G}_i \in \mathcal{G}$ which can not be distinguished by an index I. In Table 1, we see the results when using the novel descriptors based on the underlying information polynomial $P^I_{G,M}(x)$. In Table 2, we summarize the results when using the same descriptors (these are $I^1_{A,G}, I^2_{A,G}, I^3_{A,G}, I^4_{A,G}$) based on the ordinary characteristic polynomial $P^I_{G,A}(x) := \det(A - xE), A$ is the adjacency matrix of G. For both polynomials $P^I_{G,M}(x)$ and $P^I_{G,A}(x)$, we observe that the sums of the square roots of the moduli of their real zeros ($I^1_{M,G}$ and $I^1_{A,G}$) have a high discrimination power. But by evaluating the characteristic polynomial instead of the information polynomial, one can see that the sensitivity values are a little less. Further, $I^2_{M,G}$, $I^3_{M,G}$ and $I^4_{M,G}$ also do have a high discrimination power based on their sensitivity values shown in Table 1. As expected and known [41], the Wiener Index [42] and Randić Index [43] do possess a relatively high degree of degeneracy leading to low sensitivities. However, it is known that, for instance, information-theoretic

Table 1. Calculation of sensitivity index $S(I^j_{M,G}), j = 1, \ldots, 4$ for two chemical databases.

Descriptor I	S(I) for MS 2265	S(I) for AG 3982
$I^1_{M,G}$	0.999117	0.998995
$I^2_{M,G}$	0.997351	0.994224
$I^3_{M,G}$	0.998234	0.997489
$I^4_{M,G}$	0.988521	0.976645
W	0.067108	0.343046
R	0.232546	0.402562
I_D	0.859602	0.938724
I^W_D	0.883885	0.947513

doi:10.1371/journal.pone.0011393.t001

measures [14,25] do often have a high discrimination power. For comparing such measures with ours, we also calculated the sensitivities of the two entropic measures

$$I_D(G) := -\frac{1}{|V|} \log\left(\frac{1}{|V|}\right) - \sum_{i=1}^{\rho(G)} \frac{2k_i}{|V|^2} \log\left(\frac{2k_i}{|V|^2}\right), \quad (24)$$

$$I^W_D(G) := -\sum_{i=1}^{\rho(G)} \frac{ik_i}{W} \log\left(\frac{i}{W}\right), \quad (25)$$

which were developed by Bonchev [25]. Here, k_i stands for the occurrence of the distance value i in the distance matrix D. We see in Table 1 that the resulting sensitivities are much better than the ones obtained by calculating W and R.

Now, we evaluate the discrimination power of the polynomial-based descriptors. Corresponding to the fact that the characteristic polynomial of a graph is degenerated [41] (i.e., many non-isomorphic graphs have the same spectra), it is not surprising that the sensitivity values for $I^2_{A,G}$, $I^3_{A,G}$ and $I^4_{A,G}$ are much less (for both MS 2265 and AG 3982) than in case of considering the information polynomial. In fact, the discrimination power of $I^3_{M,G}, I^4_{M,G}$ when taking the characteristic polynomial into account (i.e., they become to $I^3_{A,G}, I^4_{A,G}$) is very little. Finally, we also remark that $I^j_{M,G}, j = 1, \ldots, 4$ are less degenerated than I_D and I^W_D.

Also, Table 3 shows some characteristics of the spectra concerning MS 2265 and AG 3982. For instance, $|z \in Z_{M(G)}(x) < 0|$ stands for the number of zeros (eigenvalues) of the information polynomial less than zero. τ stands for the minimal eigenvalue and σ for the maximal one. The table gives information

Table 2. Calculation of sensitivity index $S(I^j_{A,G}), j = 1, \ldots, 4$ for two chemical databases.

Descriptor I	S(I) for MS 2265	S(I) for AG 3982
$I^1_{A,G}$	0.928918	0.964591
$I^2_{A,G}$	0.699338	0.834003
$I^3_{A,G}$	0.000883	0.000883
$I^4_{A,G}$	0.01557	0.01557

doi:10.1371/journal.pone.0011393.t002

Table 3. Characteristics of the spectra concerning MS 2265 and AG 3982.

	τ	σ	$|z \in Z_{M(G)}(x) < 0|$	$|z \in Z_{M(G)}(x) > 0|$	**% < 0**	**% > 0**
MS 2265	−0.02	18.99	7798.00	21419.00	0.25	0.69
AG 3982	−0.02	109.00	20539.00	56159.00	0.25	0.70

doi:10.1371/journal.pone.0011393.t003

about the distribution of the zeros with respect to the used databases. For both MS 2265 and AG 3982, we get the result that approximately 70% of the eigenvalues are positive, 25% are negative.

As conclusive remarks, the obtained results (in particular, see Table 1) show that the spectrum of the information polynomial seems to contain useful information for defining special molecular descriptors. In particular, we infer that the information polynomials of our chemical graphs are less degenerated than the corresponding characteristic polynomials. From this, we also conclude that some of the novel measures $(I_{M,G}^j)$, $j = 1, \ldots, 4$) are suitable for characterizing (large) graphs structurally because they encode structural information uniquely.

Summary and Outlook

In this paper, we introduced a novel graph polynomial and derived some descriptors by using the underlying zero distribution. We summarize the main findings of our paper and some future ideas as follows:

- We started from the idea to use the probability distribution $P_G(V)$ (see Equation (5)) possessing n vertex probabilities which depend on an information functional [26] f. Instead of using $P_G(V)$ to derive entropic measures for graphs [26,27], this probability distribution served as starting point to derive a graph-theoretical matrix finally leading to the definition of the information polynomial.

- So far, spectra of graphs have already been used to characterize chemical graphs [41,44]. However, it is known that graph spectra (with respect to the ordinary characteristic polynomial) are degenerated, that means, many non-isomorphic graphs do have the same characteristic polynomials and spectra. For this reason, we investigated the ability of the information polynomial and the derived descriptors to discriminate graphs structurally. In particular, we found that the descriptors based on the information polynomial do have a high discrimination power by evaluating the sensitivity index due to Konstantinova [32]. But we emphasize that the achieved results depend on the used datasets (see Section 'Datasets'). As future work, we want to further develop spectral descriptors and evaluate them by using large sets of chemical and bio-chemical structures.

- It is well-known that various topological descriptors have been successfully used for structure-oriented drug design [45–47]. Typically, the descriptors are used to predict a biological or physico-chemical property by taking the structure of the underlying molecule into account. The next step in this direction would be to apply our novel descriptors to further datasets for evaluating their usefulness for solving QSAR/QSPR problems. As a strong point, we have proven that our

novel indices encode structural information uniquely that is generally a desirable property of a molecular descriptor. This could indicate their usefulness when applying the measures in drug design, e.g., for classifying chemical structures. However, we note that also degenerated measures were found to be successful for QSAR modeling, see [48].

- There is a considerable body of literature dealing with examining the zeros of graph polynomials [49–51]. For example, Woodall [51] examined the zeros of chromatic and flow polynomials and determined zero-free regions thereof. When giving [50] a survey on counting hypercubes, Kovše [50] also sketched some results concerning the zeros of cube polynomials. Jackson [49] surveyed results and conjectures dealing with the zero distribution of chromatic and flow polynomials of graphs, and characteristic polynomials of matroids. Note that a recent overview on such results was given by Ellis-Monaghan et al. [6]. As future work, we want to study similar questions concerning the information polynomial. This would involve examining problems such as:

  - Deriving special estimations for the largest positive zero of $P_{G,M}^I(x)$.
  - Deriving special bounds (leading to intervals containing the real zeros of $P_{G,M}^I(x)$) for the moduli of the real zeros of $P_{G,M}^I(x)$.
  - Exploring the location of the zeros for similar types of graph polynomials. Particularly, we want to derive special bounds for special graph classes.

- As mentioned, the largest positive eigenvalue of trees has been used as a measures for branching [13] where Bonchev [52] gave an overview on the concept of branching and measures to quantify it [52]. In the future, investigate the largest positive eigenvalue of our information polynomial in depth and compare the results with other graph polynomials.

Author Contributions

Conceived and designed the experiments: MD LAJM AG. Performed the experiments: MD LAJM AG. Analyzed the data: MD LAJM AG. Contributed reagents/materials/analysis tools: MD LAJM AG. Wrote the paper: MD LAJM AG.

References

1. Bang-Jensen J, Gutin G (2002) Digraphs. Theory, Algorithms and Applications. London, Berlin, Heidelberg: Springer.

2. Bonchev D, Rouvray DH (1991) Chemical Graph Theory. Introduction and Fundamentals. New York, NY, USA: Abacus Press.

3. Dong FM, Koh KM, Teo KL (2005) Chromatic Polynomials and Chromaticity of Graphs World Scientific Publishing Company.

4. Gutman I (1991) Polynomials in graph theory. In: Bonchev D, Rouvray DH, eds. Chemical Graph Theory Introduction and Fundamentals, Abacus Press. pp 133–176.

5. Pachter L, Sturmfels B (2005) Algebraic Statistics for Computational Biology Cambridge University Press.

6. Ellis-Monaghan JA, Merino C (2010, to appear) Graph polynomials and their applications i: The tutte polynomial. In: Dehmer M, ed. Structural Analysis of Complex Networks. Boston/Basel: Birkhäuser.

7. Bonchev D, Trinajstić N (1977) Information theory, distance matrix and molecular branching. J Chem Phys 67: 4517–4533.

8. Diudea MV, Gutman I, Jäntschi L (2001) Molecular Topology. Nova Publishing. New York, NY, USA.

9. Ivanciuc O, Ivanciuc T, Diudea MV (1999) Polynomials and spectra of molecular graphs. Roumanian Chemical Quarterly Reviews 7: 41–67.

10. Cvetkovic DM, Doob M, Sachs H (1997) Spectra of Graphs. Theory and Application Academic Press.

11. Hosoya H (1988) On some counting polynomials. Discrete Applied Mathematics 19: 239–257.

12. Randić M, Vracko M, Novic M (2001) Eigenvalues as molecular descriptors. In: Diudea MV, ed. QSPR/QSAR Studies by Molecular Descriptors, Nova Publishing. pp 93–120.

13. Lovász L, Pelikán J (1973) On the eigenvalues of trees. Periodica Mathematica Hungarica 3: 175–182.

14. Todeschini R, Consonni V, Mannhold R (2002) Handbook of Molecular Descriptors. Weinheim, Germany: Wiley-VCH.

15. Robles-Kelly A, Hancock R (2005) Graph edit distance from spectral seriation. IEEE Transactions on Pattern Analysis and Machine Intelligence 27: 365–378.

16. Mason O, Verwoerd M (2007) Graph theory and networks in biology. IET Systems Biology 1: 89–119.

17. Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences Cambridge University Press.

18. Estrada E (2002) Characterization of the folding degree of proteins. Bioinformatics 18: 697–704.

19. Chen SJ, Dill KA (1996) Symmetries in proteins: A knot theory approach. J Chem Phys 104: 5964–5973.

20. Emmert-Streib F (2006) Algorithmic computation of knot polynomials of secondary structure elements of proteins. Journal of Computational Biology 13: 1503–1512.

21. Cover TM, Thomas JA (2006) Elements of Information Theory. Wiley Series in Telecommunications and Signal Processing Wiley & Sons.

22. Mignotte M, Stefanescu D (1999) Polynomials: An Algorithmic Approach (Discrete Mathematics and Theoretical Computer Science Springer.

23. Fukui M, Kamae T (1984) Information polynomials of graphs. In: Akiyama J, ed. Number Theory and Combinatorics, World Scientific Publishing. pp 97–104.

24. Kamae R (1983) Information of relative pairwise comparisons. Journal of Mathematical Analysis and Applications 92: 355–373.

25. Bonchev D (1983) Information Theoretic Indices for Characterization of Chemical Structures. Chichester: Research Studies Press.

26. Dehmer M, Emmert-Streib F (2008) Structural information content of networks: Graph entropy based on local vertex functionals. Comput Biol Chem 32: 131–138.

27. Dehmer M, Varmuza K, Borgert S, Emmert-Streib F (2009) On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures. J Chem Inf Model 49: 1655–1663.

28. Mowshowitz A (1968) Entropy and the complexity of the graphs I: An index of the relative complexity of a graph. Bull Math Biophys 30: 175–204.

29. Shannon CE, Weaver W (1997) The Mathematical Theory of Communication. Urbana, IL, USA: University of Illinois Press.

30. Bretscher O (2004) Linear Algebra with Applications Prentice Hall; 3rd edition.

31. Devillers J, Balaban AT (1999) Topological Indices and Related Descriptors in QSAR and QSPR. Amsterdam, The Netherlands: Gordon and Breach Science Publishers.

32. Konstantinova EV (1996) The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems. J Chem Inf Comput Sci 36: 54–57.

33. Konstantinova EV (2006) On some applications of information indices in chemical graph theory. In: Ahlswede R, Bäumer L, Cai N, Aydinian H, Blinovsky V, et al. (2006) General Theory of Information Transfer and Combinatorics, Springer, Lecture Notes of Computer Science. pp 831–852.

34. Todeschini R, Cazar R, Collina E (1992) The chemical meaning of topological indices. Chemometrics and Intelligent Laboratory Systems 15: 51–59.

35. Schwaighofer A, Schroeter T, Mika S, Hansen K, Laak AT, et al. (2008) A probabilistic approach to classifying metabolic stability. J Chem Inf Model 48: 785–796.

36. Hansen K, Mika S, Schroeter T, Sutter A, Laak AT, et al. (2009) A benchmark data set for in silico prediction of ames mutagenicity. J Chem Inf Model.

37. Scsibrany H, Varmuza K (2004) Software SubMat. www.lcm.tuwien.ac.at. Vienna University of Technology, Institute of Chemical Engineering, Laboratory for Chemometrics, Austria.

38. Stein SE (1998) NIST, Mass spectral database 98. www.nist.gov/srd/nist1a.htm. National Institute of Standards and Technology, Gaithersburg, MD, USA.

39. (2008) R, software, a language and environment for statistical computing. www.r-project.org. R Development Core Team, Foundation for Statistical Computing, Vienna, Austria.

40. Gasteiger J, Engel T (2003) Chemoinformatics - A Textbook. Weinheim, Germany: Wiley VCH.

41. Ivanciuc O, Balaban AT (1999) The graph description of chemical structures. In: Devillers J, Balaban AT, eds. Topological Indices and Related Descriptors in QSAR and QSPAR, Gordon and Breach Science Publishers 59–167, Amsterdam, The Netherlands, .

42. Wiener H (1947) Structural determination of paraffin boiling points. Journal of the American Chemical Society 69: 17–20.

43. Randić M (1975) On characterization of molecular branching. J Amer Chem Soc 97: 6609–6615.

44. Trinajstić N (1992) Chemical Graph Theory. Boca Raton, FL, USA: CRC Press.

45. Bajorath J (2004) Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery. Totowa, NJ, USA: Methods in Molecular Biology. Humana Press.

46. Basak SC (1999) Information-theoretic indices of neighborhood complexity and their applications. In: Devillers J, Balaban AT, eds. Topological Indices and Related Descriptors in QSAR and QSPAR, Gordon and Breach Science Publishers 563–595, Amsterdam, The Netherlands, .

47. Basak SC, Nikolić S, Trinajstić N, Amic D, Beslo D (2000) QSPR modeling: Graph connectivity indices versus line graph connectivity indices. J Chem Inf Comput Sci 40: 927–933.

48. Todeschini R, Consonni V, Maiocchi A (1998) The k correlation index: theory development and its applications in chemometrics. Chemometrics and Intelligent Laboratory Systems 46: 13–29.

49. Jackson B (2003) Zeros of chromatic and flow polynomials of graphs. Journal of Geometry 76: 95–109.

50. Kovše M (2009) Counting cubes in median graphs and related problems. In: Analysis of Complex Networks: From Biology to Linguistic, Wiley VCH 323–350, Weinheim, Germany, .

51. Woodall D (1992) A zero-free interval for chromatic polynomials. Discrete Mathematics 101: 333–341.

52. Bonchev D (1995) Topological order in molecules 1. molecular branching revisited. Journal of Molecular Structure: THEOCHEM 336: 137–156.